



Trực quan hóa dữ liệu



Nội dung

- ❖ Chọn biểu đồ phù hợp cho loại biến phù hợp hoặc tập các biến
- ❖ Hiểu mối nguy hiểm của đề xuất phụ thuộc khi sự phụ thuộc không tồn tại
- ❖ Thảo luận ưu nhược điểm của thêm màu vào hiển thị dữ liệu trực quan hóa dữ liệu
- ❖ Hạn chế trang trí biểu đồ với các hiệu ứng



TẦM QUAN TRỌNG HIỂN THỊ DỮ LIỆU

- ❖ Trực quan hóa hiển thị dữ liệu giúp khám phá các mẫu trong dữ liệu và giúp hiểu các xu hướng không dễ thấy ở các bảng tổng hợp.
- ❖ Việc hiển thị dữ liệu một cách hợp lý cũng giúp tránh quá tải thông tin đối với người phân tích dữ liệu.
- ❖ Sự ưu việt của việc sử dụng các biểu đồ trong phân tích dữ liệu so với các bảng tổng hợp đã được chứng minh bằng thực nghiệm người dùng phân tích dữ liệu và ra quyết định.



BIỂU DIỄN TRỰC QUAN MỘT BIẾN

- ❖ Các biểu đồ này hiển thị một tập các giá trị của một biến cụ thể.
- ❖ Trong các trường hợp các biến được đo trong các thang đo thứ tự, khoảng và tỷ lệ, các thang đo này đủ thông tin để có thể được vẽ các giá trị dữ liệu theo một trục.
- ❖ Các dữ liệu riêng biệt và tổng hợp có thể được sử dụng để vẽ các biến trên các trục.
- ❖ Một trục biểu diễn một chiều của dữ liệu.



BIỂU DIỄN TRỰC QUAN MỘT BIẾN

Biểu đồ đường → 

Biểu đồ hộp → 

Biểu đồ
thang chấm → 



BIỂU ĐỒ THANG CHẤM

- ❖ Biểu diễn thang chấm: mỗi giá trị được biểu diễn bởi một chấm trên thang của biến → cho ta hiểu biết về phân bố của các giá trị dữ liệu riêng biệt trên toàn dải giá trị.
- ❖ Thang chấm bắt đầu mất giá trị khi có nhiều điểm dữ liệu và phân bố các giá trị dọc trục trở lên không rõ ràng → Biểu diễn các phép đo tổng hợp thống kê sẽ hiệu quả hơn.



BIỂU ĐỒ HỘP

- ❖ Biểu đồ hộp: bắt đầu ở từ phân vị thứ nhất và kết thúc ở tứ phân vị thứ ba.
- ❖ Dải trong hộp là dải liên tứ phân vị (một nửa các giá trị).
- ❖ Đường thẳng bên trong hộp biểu diễn giá trị trung vị.
- ❖ Các đường bên ngoài hộp gọi là các viền phân nhánh tới giá trị thấp nhất và cao nhất.



BIỂU ĐỒ ĐƯỜNG

- ❖ Biểu đồ đường: biểu diễn phép đo thống kê giống như biểu đồ hộp chỉ khác ở chỗ giá trị trung vị được biểu diễn bởi một dấu chấm.
- ❖ Khoảng trắng giữa các đường và dấu chấm biểu diễn dải liên tứ phân vị.

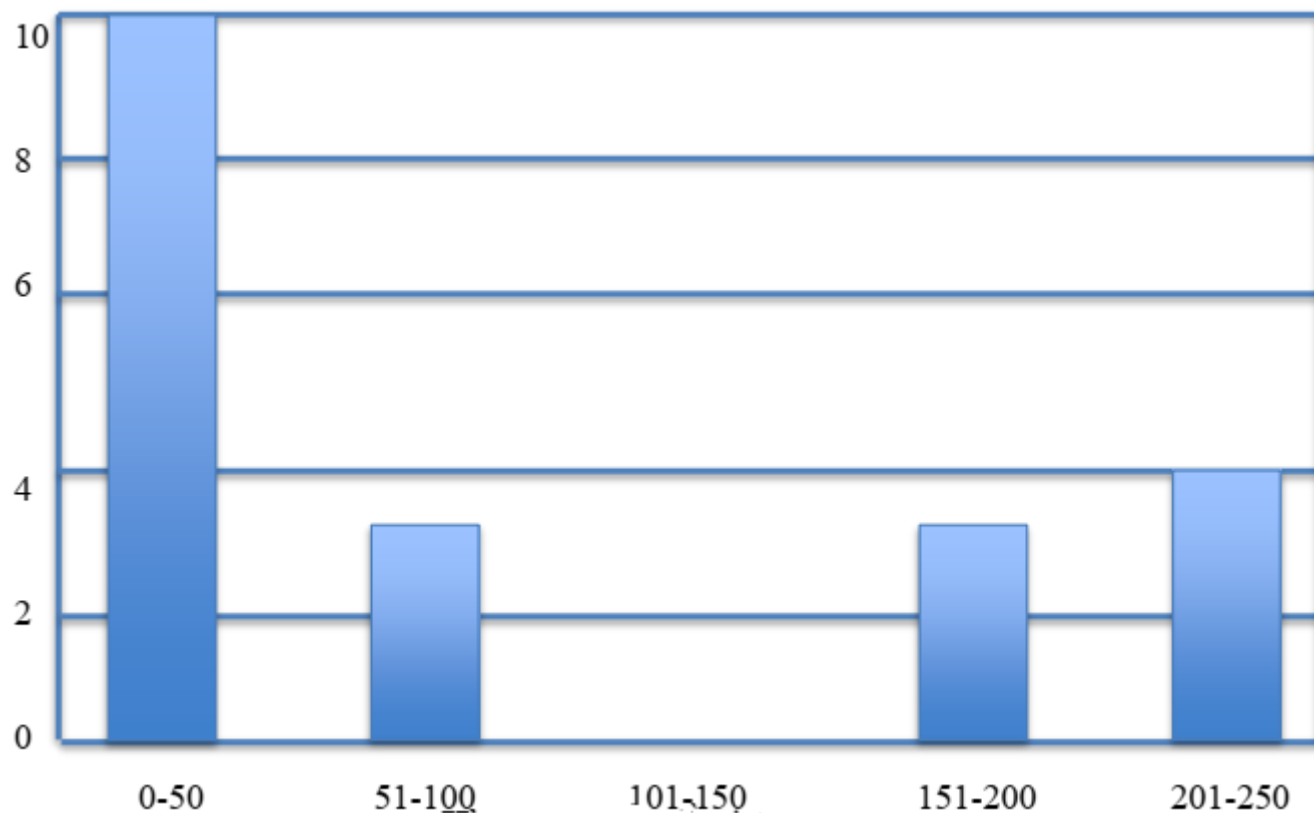


BIỂU ĐỒ HISTOGRAM

- ❖ Bảng tần số là một dạng bảng tóm tắt đặc biệt cung cấp các phép đo tổng hợp cho một biến.
- ❖ Loại biểu đồ được chọn để hiển thị các bảng tần số phụ thuộc vào thang đo của biến.
Biểu đồ histogram: mô tả bảng tần số cho các biến thang thứ tự, khoảng hoặc tỷ lệ
- ❖ Trục hoành biểu diễn các khoảng giá trị và trục tung biểu diễn tần số.



BIỂU ĐỒ HISTOGRAM



Ví dụ biểu đồ histogram biểu diễn tổng số các sản phẩm trên các dải giá trị lợi tức.

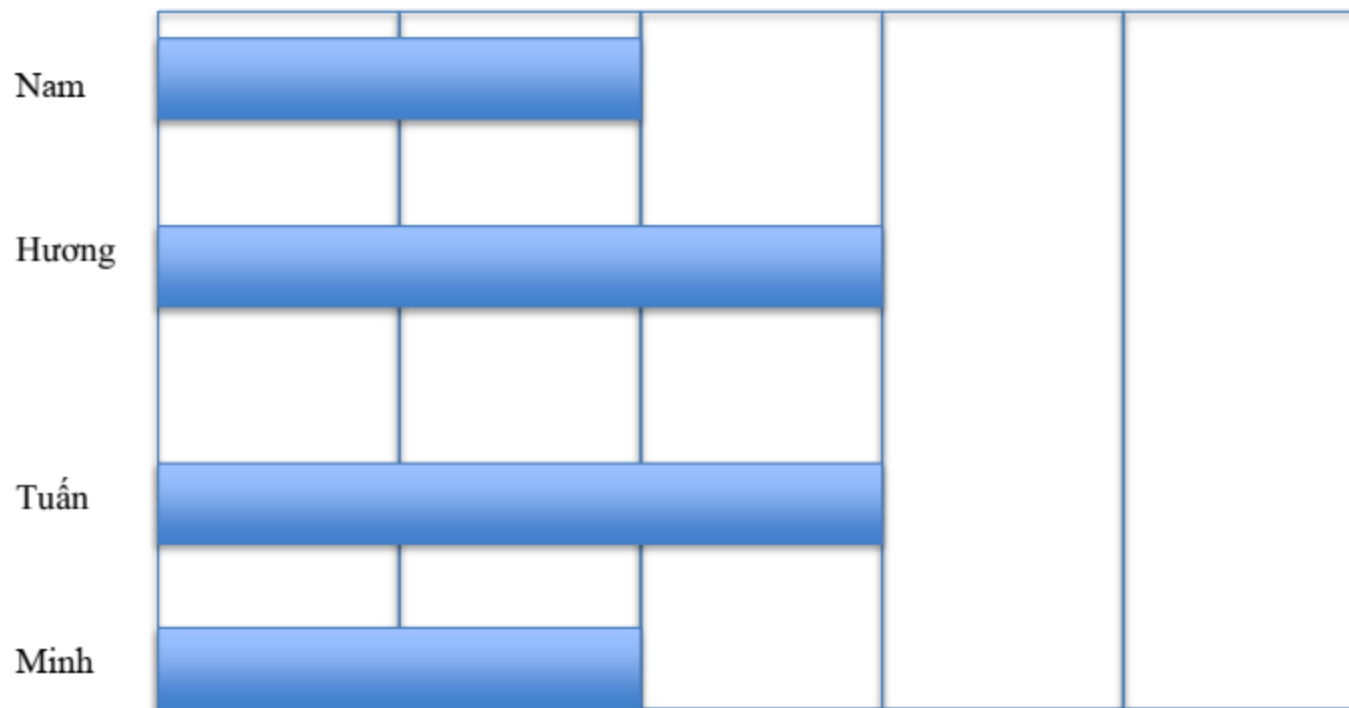


Biểu đồ thanh (bar chart)

- ❖ Biểu đồ bar chart: biểu diễn bảng tần số. Một biểu đồ thanh giống như biểu đồ histogram nhưng các giá trị danh định không biểu diễn trình tự nào cả.
- ❖ Biểu đồ thanh thường được miêu tả theo chiều ngang.
- ❖ Ví dụ: số lần mỗi nhân viên kinh doanh thực hiện thành công một đơn hàng



Biểu đồ thanh (bar chart)





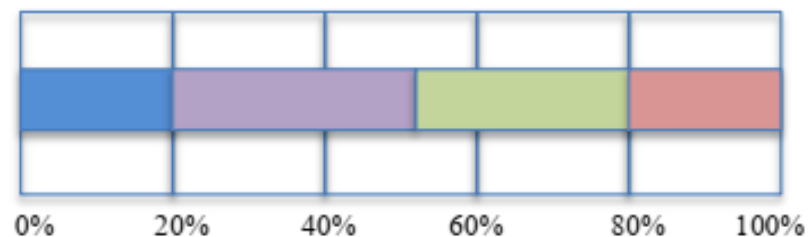
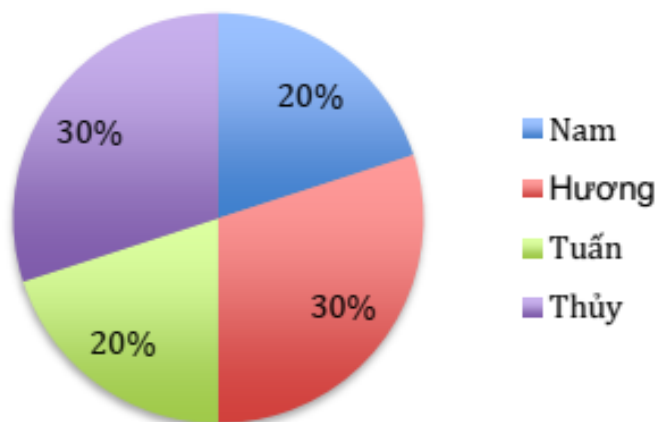
Biểu đồ tròn và chồng xếp

- ❖ Nếu chúng ta quan tâm đến tần số tương đối hơn là tần số tuyệt đối, một biến có thể được biểu đồ hóa với các biểu đồ dạng tròn (pie chart) hoặc biểu đồ thanh chồng xếp (stacked bar chart).
- ❖ Loại biểu đồ này được nhấn mạnh biểu diễn các tỷ lệ phần trăm các giá trị dữ liệu.



Biểu đồ tròn và chồng xếp

Bán hàng





BIỂU DIỄN TRỰC QUAN HIỂN THỊ HAI BIẾN

- ❖ Cung cấp dữ liệu tổng hợp của một biến được nhóm theo các giá trị của biến khác.
- ❖ Dữ liệu tổng hợp thường biểu diễn ở thang đo khoảng hoặc tỷ lệ và các biến nhóm thường ở thang danh định.
- ❖ Bảng tổng hợp giống như bảng tần số có thể được biểu đồ hóa sử dụng các biểu đồ thanh và biểu đồ dạng tròn.
- ❖ Các bảng cross-tab và bảng pivot là các bảng tóm tắt được tách làm ma trận hai nhân hai.

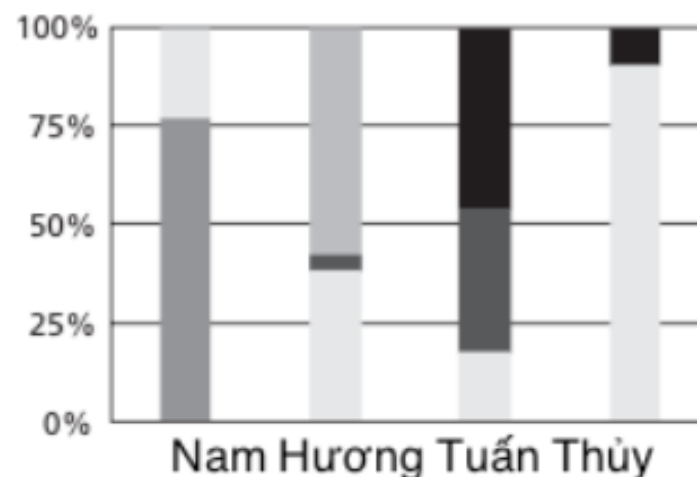
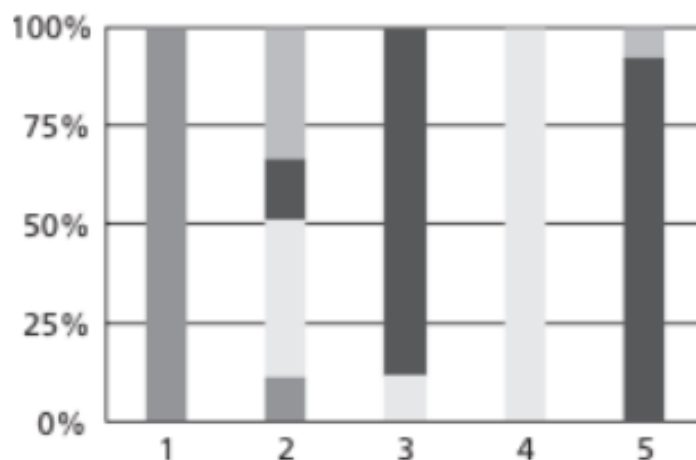
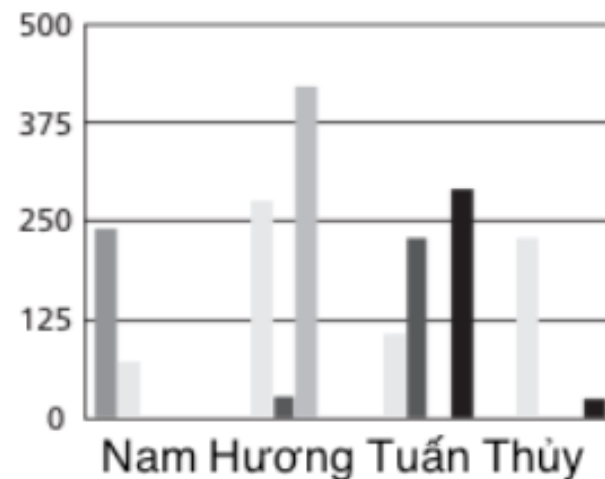
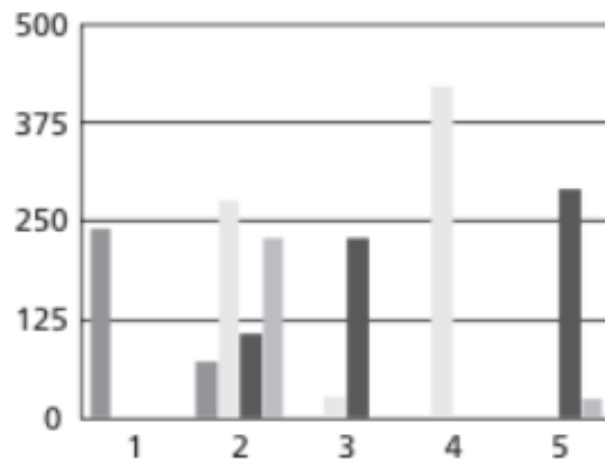


BIỂU DIỄN TRỰC QUAN HIỂN THỊ HAI BIẾN

- ❖ Biểu diễn cùng các giá trị tổng hợp được chia làm hai chiều.
- ❖ Cách thông thường để biểu đồ hóa các bảng cross-tab sử dụng các biến thể đặt cạnh nhau (side-by-side) của biểu đồ đơn biến.
- ❖ Dùng biểu đồ đơn biến để biểu đồ hóa bảng cross-tab



Biểu đồ đơn biến biểu đồ hóa bảng cross-tab





Biểu đồ đơn biến biểu đồ hóa bảng cross-tab

- ❖ Biểu đồ góc trên bên trái: đối chiếu các nhân viên kinh doanh với nhau và sắp xếp thứ tự theo tuần.
- ❖ Biểu đồ góc trên bên phải đối chiếu các tuần và sắp xếp theo nhân viên kinh doanh.
- ❖ Hai biểu đồ bên dưới trong hình là các biến thể side-by-side của biểu đồ chồng xếp tập trung vào các tần số tương đối các giá trị dữ liệu



Biểu đồ hai biến

- ❖ Khai báo các biến là biến độc lập và các thuộc tính khác là các biến phụ thuộc.
- ❖ Biến độc lập là biến biểu diễn nguyên nhân, biến phụ thuộc biểu diễn kết quả → đã định nghĩa chiều hiệp biến.
- ❖ Nếu biến A là biến độc lập và biến B là biến phụ thuộc nghĩa là thay đổi biến A sẽ dẫn đến thay đổi biến B nhưng không đúng theo chiều ngược lại.
- ❖ Biến độc lập được biểu diễn trên trục hoành còn biến phụ thuộc sẽ được biểu diễn trên trục tung.



Biểu đồ hai biến

Entity	Var1	Var2	Var3	Var4
1	60	21	40	8
2	35	8	33	5
3	78	30	60	14
4	71	15	51	10
5	53	17	43	10
6	57	9	35	6
7	25	3	25	7
8	45	13	38	11
9	49	18	43	13
10	43	13	29	8
11	39	7	32	7
12	51	7	35	5
13	44	8	35	5
14	75	17	54	18
15	38	8	46	9
16	83	29	49	13
17	60	16	52	12
18	41	4	27	5
19	40	13	35	7
20	58	11	47	8

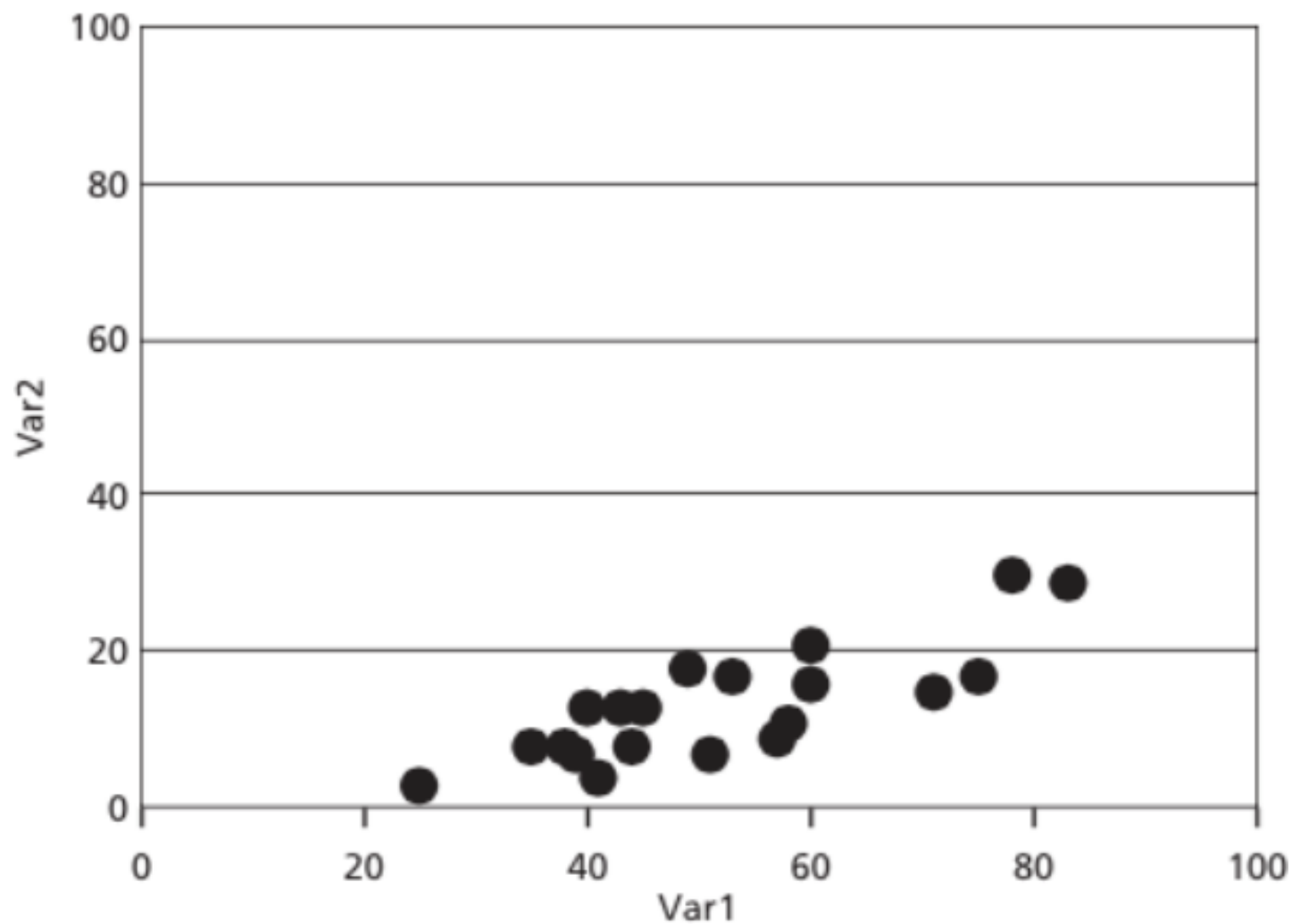


Biểu đồ hai biến

- ❖ Biểu đồ phân bố là mở rộng nhị biến của biểu đồ chấm ở đó mỗi chấm biểu diễn hai giá trị, một trên trục hoành và một trên trục tung.
- ❖ Biểu đồ hóa một biểu đồ phân bố không hàm ý là các biến hiệp biến hoặc tồn tại sự phụ thuộc giữa chúng.
- ❖ Mục đích của biểu đồ phân bố là để tìm xem có tồn tại hiệp biến giữa hai biến hay không.
- ❖ Biểu đồ phân bố cho hai biến Var1 và Var2 → Có sự hiệp biến tuyến tính giữa hai biến Var1 và Var2.



Biểu đồ hai biến



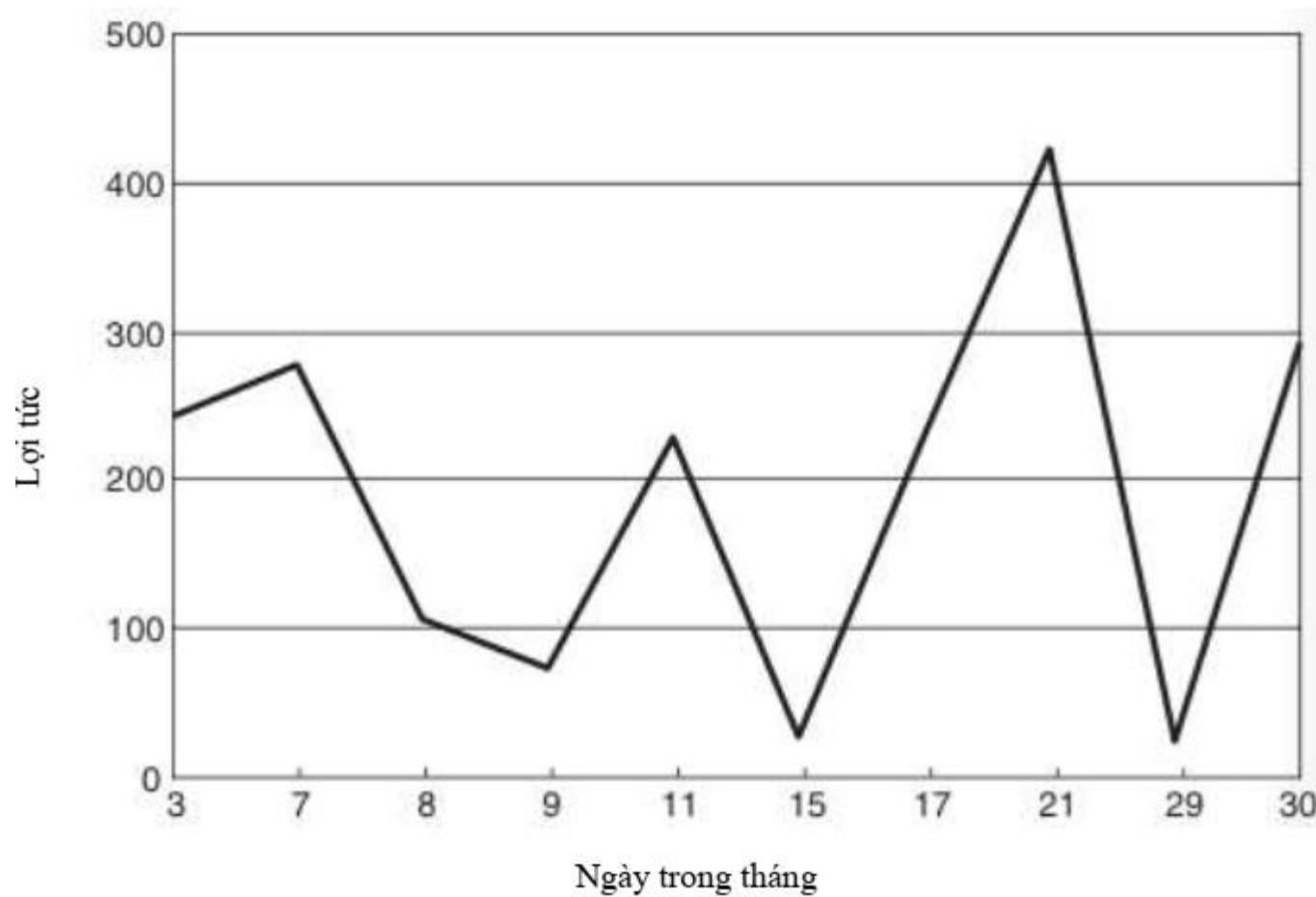


Biểu đồ đường

- ❖ Biểu đồ đường là một loại đặc biệt của biểu đồ phân bố ở đó một đường được vẽ nối giữa các điểm.
- ❖ Đường gợi ý xu hướng của biến và hàm ý rằng các giá trị trên đường có thể được suy ra.
- ❖ Một loại biểu đồ đường là biểu đồ chuỗi thời gian sử dụng một biến chuỗi thời gian được hiển thị trên trục hoành và một số biến khác hiển thị trên trục tung.
- ❖ Ví dụ: một biểu đồ đường biểu diễn thay đổi lợi tức theo thời gian.



Biểu đồ đường



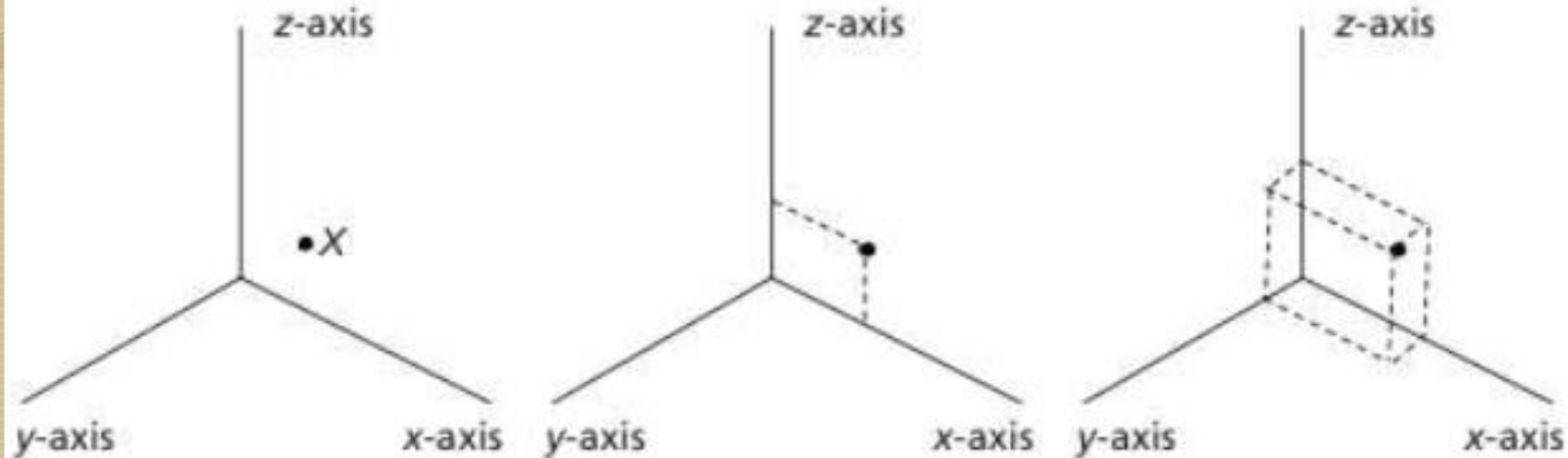


Biểu diễn trực quan hiển thị ba hoặc nhiều biến

- ❖ Sử dụng các biểu đồ phân bố 3 chiều và biểu đồ đường 3 chiều (trục X , Y , Z).
- ❖ Khi biểu diễn trong không gian 3 chiều, cần có các đường hỗ trợ để tránh nhầm lẫn thị giác về vị trí của các điểm.
- ❖ Tránh nhầm lẫn thị giác về vị trí của một điểm trong không gian



Nhầm lẫn thị giác vị trí biểu đồ 3 chiều

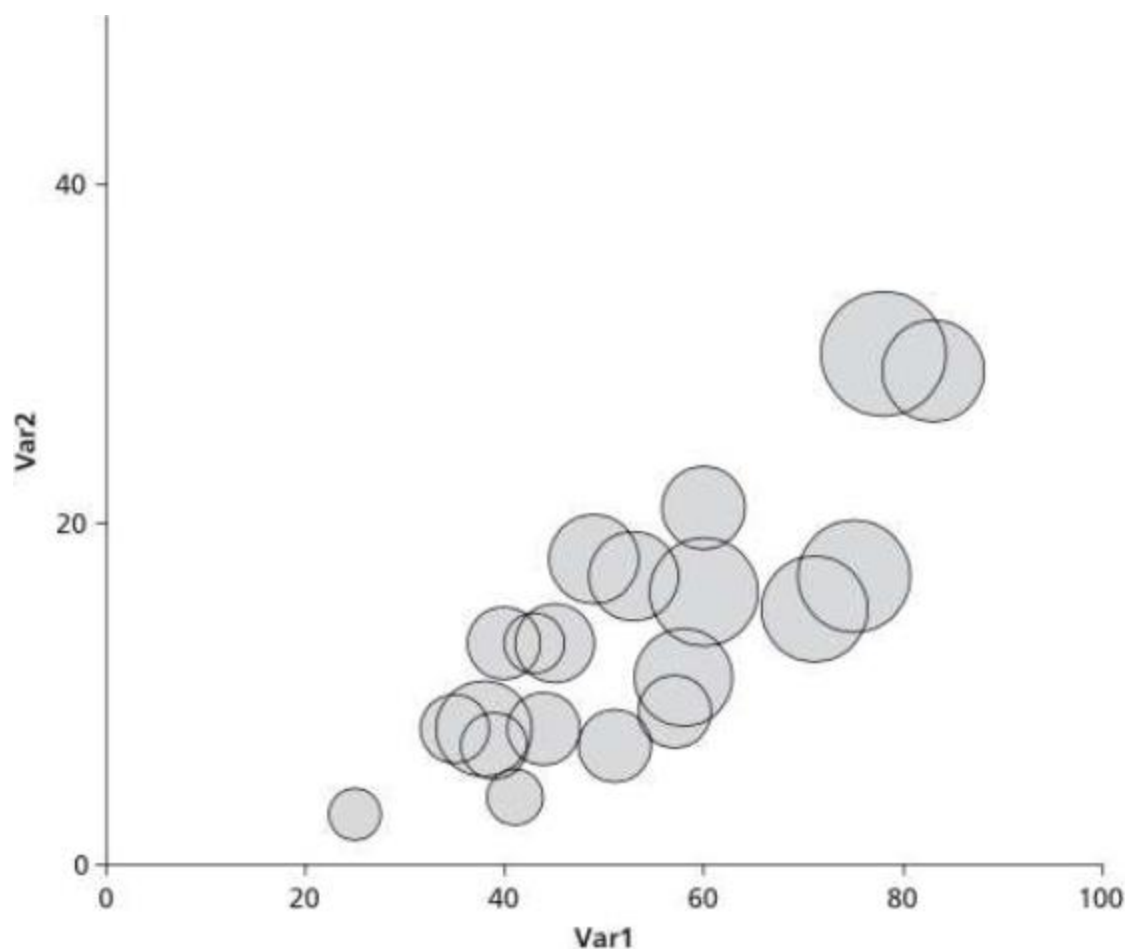


Ví dụ về nhầm lẫn thị giác về vị trí trên biểu đồ phân bố 3 chiều.



Biểu đồ bong bóng (bubble chart)

- ❖ Biểu đồ bong bóng giống như biểu đồ phân bố nhị biến chỉ khác ở chỗ chiều thứ 3 được biểu diễn bởi đường kính của chấm.





Biểu đồ bong bóng (bubble chart)

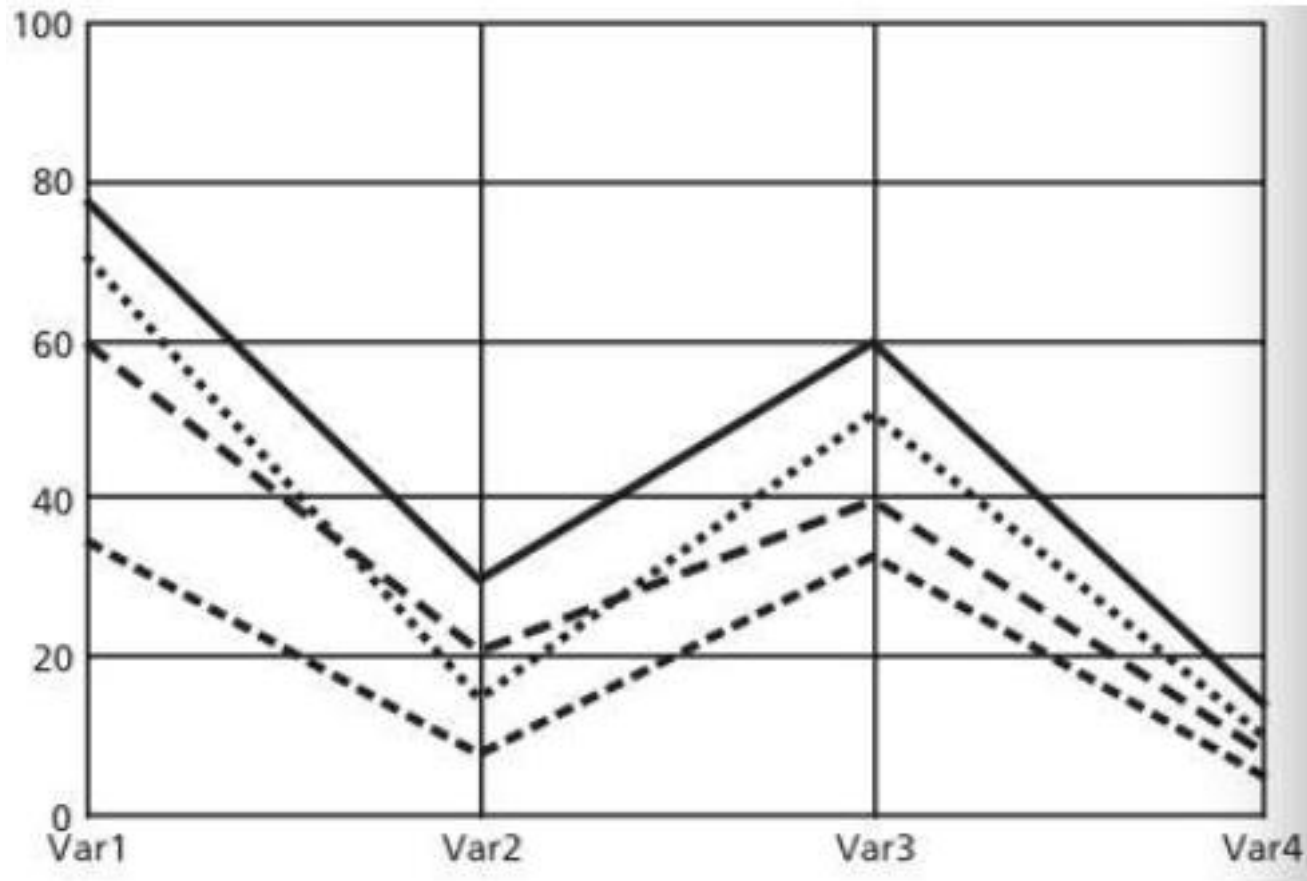
- ❖ Để biểu diễn bốn chiều, biểu đồ bong bóng có thể được sử dụng ở đó chiều thứ 4 được biểu diễn bởi khuynh độ màu
- ❖ Màu của các bong bóng trong biểu đồ bong bóng với giá trị tương ứng biểu diễn biến thứ tư.
- ❖ Vấn đề đối với biểu đồ bong bóng tô màu là rất khó thấy các mối quan hệ giữa các chiều liên quan.



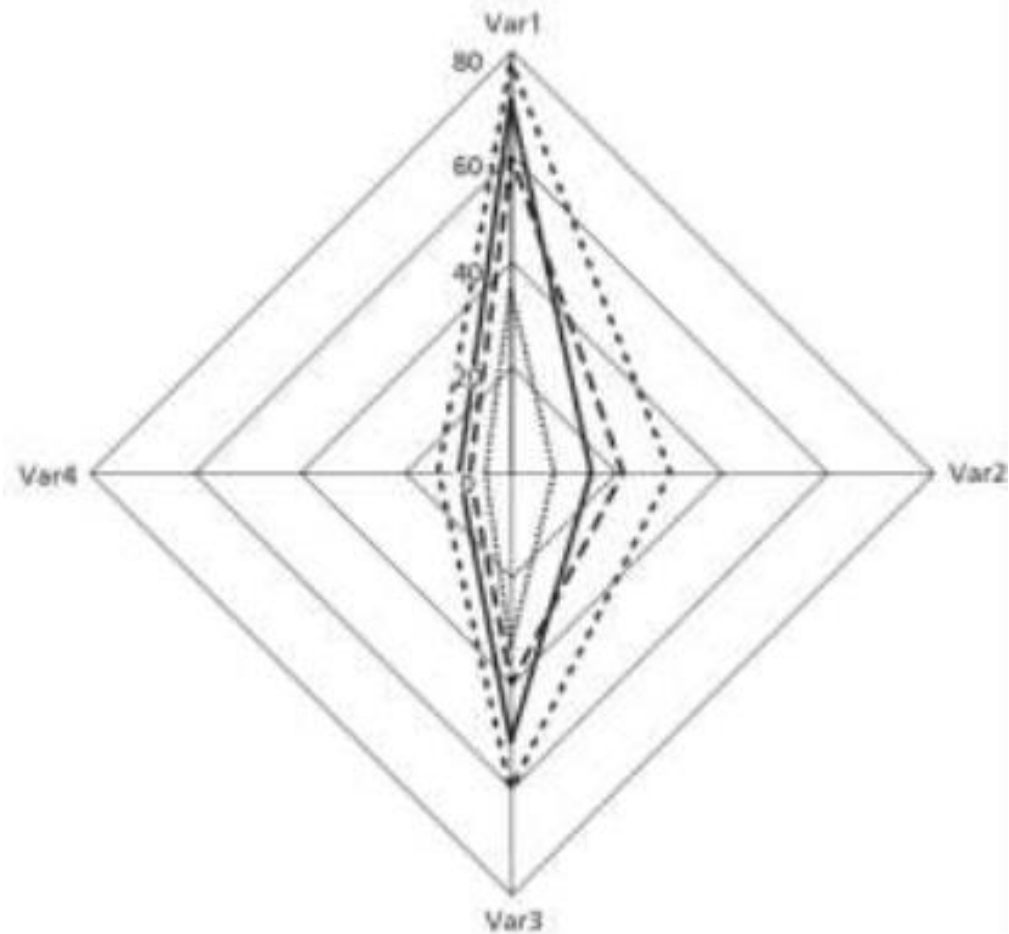
Biểu đồ cho dữ liệu đa chiều

- ❖ Dữ liệu đa chiều với năm hay sáu chiều hoặc nhiều chiều hơn rất khó biểu diễn trực quan.
- ❖ Kỹ thuật phổ biến nhất là biểu đồ phối hợp song song: các trục được đặt song song với nhau cho mỗi biến và mỗi thể hiện dữ liệu một đường thẳng được vẽ nối các giá trị trên mỗi trục.
- ❖ Khi các trục không được đặt song song với nhau mà theo vòng tròn thì biểu đồ được gọi là biểu đồ ra đa hoặc biểu đồ sao.

Biểu đồ phối hợp song song



Biểu đồ ra đa (biểu đồ sao)





Biểu đồ động

- ❖ Biểu đồ động là các biểu đồ thay đổi diện mạo khi tương tác với người dùng.
- ❖ Sử dụng biểu đồ động là cách hiệu quả để biểu diễn trực quan các xu hướng và tiết kiệm không gian thay vì sử dụng một loạt các biểu đồ tĩnh.
- ❖ Biểu đồ động chỉ có thể được cài đặt trong các hệ thống thông tin được máy tính hóa và không thể sử dụng trong các báo cáo quản lý truyền thống dựa trên bản giấy.



Kỹ thuật cọ vẽ (brushing)

- ❖ Kỹ thuật cọ vẽ (brushing) được mô tả bởi Becker và Cleveland năm 1987 sử dụng chuột máy tính để tương tác với biểu đồ động.
- ❖ Khi dùng chuột trở đến một chấm nào đó hoặc một khu vực trong biểu đồ thì biểu đồ sẽ thay đổi hình dạng, làm nổi bật so với các chấm khác.
- ❖ Hoặc khi di chuyển cọ vẽ trên biến chuỗi thời gian, biểu đồ sẽ hiển thị thông tin chi tiết về thời gian tại vị trí con trỏ.



Kỹ thuật chú thích tương tác

- ❖ Kỹ thuật chú thích tương tác: có quá nhiều thể hiện hoặc giá trị phân loại, chú thích có thể được sử dụng trong tương tác. Mỗi giá trị phân loại có thể được hiển thị hoặc ẩn.
- ❖ Ví dụ: trong biểu đồ phối hợp song song, chúng ta có thể hiển thị hoặc ẩn các đường của các thể hiện cụ thể để tập trung vào các thể hiện chúng ta quan tâm.



Kỹ thuật phủ (overlay)

- ❖ Kỹ thuật phủ (overlay): tổ chức các bảng và các biểu đồ thành các lớp và cho phép hiển thị hoặc ẩn các lớp.
- ❖ Một lớp có thể được xếp chồng lên lớp khác để cung cấp thêm thông tin hoặc có thể không được hiển thị để bớt thông tin được trình bày



Màu sắc và các hiệu ứng hình ảnh

- ❖ Hiệu ứng thông tin: màu sắc có thể được sử dụng để tăng thêm thông tin cho một biểu đồ. Ví dụ: giá trị 'tốt' hoặc 'xấu' trên các biểu đồ bằng các giá trị màu tương ứng xanh và đỏ.
- ❖ Biểu đồ bong bóng được sử dụng kết hợp với khuynh độ màu ở đó các giá trị màu có thể được sử dụng để biểu diễn giá trị dữ liệu ở một chiều thứ tư.
- ❖ Hiệu ứng cô lập: Một màu sử dụng để cô lập một giá trị dữ liệu từ các giá trị khác để thu hút sự chú ý đến giá trị dữ liệu đó.



Màu sắc và các hiệu ứng hình ảnh

- ❖ Hiệu ứng thẩm mỹ: có được khi thêm màu sắc gọi là sự hài hòa về màu sắc. Sự hài hòa về màu sắc là kết quả của sự kết hợp các màu cho cảm giác trình tự và cân bằng thị giác.
- ❖ Sử dụng ‘bánh xe màu’: hiển thị một số các màu lân cận nhau tạo thành một vòng tròn.
- ❖ Để có hiệu ứng hài hòa: chọn các màu tương tự nhau ở cạnh nhau phía bên phải trong bánh xe màu.
- ❖ Để có hiệu ứng tương phản: chọn các màu bù nhau là các màu nằm đối diện nhau trong bánh xe màu.



Bóp méo thị giác ở biểu đồ tròn

- ❖ Giá trị B và E là bằng nhau nhưng trong phiên bản biểu đồ 3D ấn tượng thị giác là B lớn hơn E

